EXTRACTION OF FACTS FROM TEXT

Abstract of the Invention

A fact extraction tool set ("FEX") finds and extracts targeted pieces of information from text using linguistic and pattern matching technologies, and in particular, text annotation and fact extraction. Text annotation tools break a text, such as a document, into its base tokens and annotate those tokens or patterns of tokens with orthographic, syntactic, semantic, pragmatic and other attributes. A user-defined "Annotation Configuration" controls which annotation tools are used in a given application. XML is used as the basis for representing the annotated text. A tag uncrossing tool resolves conflicting (crossed) annotation boundaries in an annotated text to produce well-formed XML from the results of the individual annotators. The fact extraction tool is a pattern matching language which is used to write scripts that find and match patterns of attributes that correspond to targeted pieces of information in the text, and extract that information.